

执行器饱和的离散时间多智能体系统有限时域一致性控制

王巍^{1,2} 王珂² 黄自鑫³ 王乐君⁴ 穆朝絮^{2,5}

摘要 针对执行器饱和的离散时间线性多智能体系统 (Multi-agent systems, MASs) 有限时域一致性控制问题, 将低增益反馈 (Low gain feedback, LGF) 方法与 Q 学习相结合, 提出采用后向时间迭代的模型无关控制方法. 首先, 将执行器饱和的有限时域一致性控制问题的求解转化为执行器饱和的单智能体有限时域最优控制问题的求解, 并证明可以通过求解修正的时变黎卡提方程 (Modified time-varying Riccati equation, MTVRE) 实现有限时域最优控制. 随后, 引入时变参数化 Q 函数 (Time-varying parameterized Q-function, TVPQF), 并提出基于 Q 学习的模型无关后向时间迭代算法, 可以更新低增益参数, 同时实现逼近求解 MTVRE. 另外, 证明所提迭代求解算法得到的 LGF 控制矩阵收敛于 MTVRE 的最优解, 也可以实现全局有限时域一致性控制. 最后, 通过仿真实验结果验证了该方法的有效性.

关键词 有限时域一致性控制, 执行器饱和, Q 函数, 模型无关, 多智能体系统

引用格式 王巍, 王珂, 黄自鑫, 王乐君, 穆朝絮. 执行器饱和的离散时间多智能体系统有限时域一致性控制. 自动化学报, 2025, 51(3): 617–630

DOI 10.16383/j.aas.c240446

CSTR 32138.14.j.aas.c240446

Finite-horizon Consensus Control of Discrete-time Multi-agent Systems With Actuator Saturation

WANG Wei^{1,2} WANG Ke² HUANG Zi-Xin³ WANG Le-Jun⁴ MU Chao-Xu^{2,5}

Abstract A model-free control method using backward-in-time iteration by combining the low gain feedback (LGF) method with Q-learning is proposed for the finite-horizon consensus control problem for discrete-time linear multi-agent systems (MASs) with actuator saturation. First, the solution of the finite-horizon consensus control problem with actuator saturation is transformed into the solution of the finite-horizon optimal control problem of single agent with actuator saturation, and it is proved that the finite-horizon optimal control can be realized by solving the modified time-varying Riccati equation (MTVRE). Then, a time-varying parameterized Q-function (TVQPF) is introduced, and a model-free backward-in-time iteration algorithm based on Q-learning is proposed to update the low gain parameter and simultaneously approximate the solution of the MTVRE. In addition, it is demonstrated that the LGF control matrix obtained by the proposed iterative solution algorithm converges to the optimal solution of the MTVRE, and the global finite-horizon consensus control can also be realized. Finally, the effectiveness of the proposed method is verified by simulation results.

Key words Finite-horizon consensus control, actuator saturation, Q-function, model-free, multi-agent systems (MASs)

Citation Wang Wei, Wang Ke, Huang Zi-Xin, Wang Le-Jun, Mu Chao-Xu. Finite-horizon consensus control of discrete-time multi-agent systems with actuator saturation. *Acta Automatica Sinica*, 2025, 51(3): 617–630

收稿日期 2024-06-30 录用日期 2024-10-28

Manuscript received June 30, 2024; accepted October 28, 2024
湖北省自然科学基金 (2023AFB561), 国家资助博士后研究人员计划 (GZB20240525) 资助

Supported by Hubei Provincial Natural Science Foundation of China (2023AFB561) and Postdoctoral Fellowship Program of CPSF (GZB20240525)

本文责任编辑 易新蕾

Recommended by Associate Editor YI Xin-Lei

1. 中南财经政法大学信息工程学院 武汉 430073 2. 天津大学电气自动化与信息工程学院 天津 300072 3. 武汉工程大学电气信息学院 武汉 430205 4. 重庆邮电大学自动化学院 重庆 400065 5. 安徽大学自主无人系统技术教育部工程研究中心 合肥 230601

1. School of Information Engineering, Zhongnan University of Economics and Law, Wuhan 430073 2. School of Electrical and Information Engineering, Tianjin University, Tianjin 300072 3. School of Electrical and Information Engineering, Wuhan Institute of Technology, Wuhan 430205 4. School of Automation, Chongqing University of Posts and Telecommunications, Chongqing

近年来, 多智能体系统 (Multi-agent systems, MASs) 分布式协同控制问题的研究取得了显著进展, 引发各个领域的广泛关注. 该研究范畴涵盖生物系统中的群体行为^[1]、分布式传感器网络技术^[2]和智能电网管理^[3]等多个方面. 一致性问题作为支撑 MASs 分布式协同控制的基础问题, 不仅在理论层面具有深远的意义, 而且在实际应用中展现出巨大的价值. 一致性控制的根本挑战在于设计高效的一致性算法或协议, 旨在确保 MASs 的所有智能体能够逐步调整其状态或输出, 最终达到相同, 即实

400065 5. Engineering Research Center of Autonomous Unmanned System Technology, Ministry of Education, Anhui University, Hefei 230601

现智能体的一致性。

目前, MASs 一致性控制的研究可以根据系统中领航者的数量划分为三类: 无领航者的一致性控制、领导-跟随一致性控制 (一个领航者) 以及包含多个领航者的一致性控制^[4-6]。到目前为止, MASs 一致性控制的研究涵盖越来越复杂的智能体动态特性和通信网络拓扑, 包括但不限于线性^[7] 或非线性 MASs^[8]、整数阶^[9] 或分数阶模型^[10]、固定^[11] 或时变拓扑^[12]、输入延迟^[13]、输入饱和^[14] 等。在上述复杂情况下, 各种适当的控制算法被提出以实现一致性控制。此外, 由于智能体的通信和计算资源有限, 基于事件触发的控制策略^[15-16] 被用于实现一致性控制, 有效减少了不必要的能源消耗。然而, 这些研究成果只能实现渐近一致性控制, 即在理论上调节时间趋于无穷。在实际应用中, 由于渐近一致性的收敛时间较长, 难以满足任务的时效性需求。

相比之下, 有限时域一致性被认为是一种更为理想的控制策略。有限时域控制不仅能够缩短闭环系统的收敛时间, 还具备更好的鲁棒性和抗干扰能力^[17]。文献 [18] 提出一种分散模型预测控制方案, 实现了一阶 MASs 的有限时域状态一致性控制。文献 [19] 采用分布式线性二次型博弈方法, 实现了离散时间二阶 MASs 的有限时域状态一致性控制。此外, 文献 [20-23] 研究离散时变 MASs 的 H_∞ 有限时域状态一致性控制问题。上述有限时域状态一致性协议的设计通常假设智能体动力学模型已知^[20-23], 或仅考虑简单的一阶^[18]、二阶^[19] 系统。然而, 一阶和二阶系统无法充分描述实际系统的动态特性, 而且在实际应用中, 系统模型信息通常是未知的或难以获取的。传统的有限时域一致性协议在系统模型未知的情况下并不适用, 难以满足实际应用的需求。

自适应动态规划 (Adaptive dynamic programming, ADP)^[24] 或强化学习 (Reinforcement learning, RL)^[25] 能够利用仿生学习机制解决系统模型未知情况下的优化控制问题^[26]。其中, 学习状态-动作值函数的 Q 学习算法^[27] 为实现无模型最优控制提供了一种有效的解决方案。近年来, 学者们利用 ADP 或 RL 算法, 通过逼近求解耦合的哈密顿-雅可比-贝尔曼 (Hamilton-Jacobi-Bellman, HJB) 方程, 以实现 MASs 的最优渐近一致性控制^[7-8, 28-31]。例如, 基于 Q 学习的算法已经应用于异构 MASs^[7-8, 28-29] 和同构 MASs^[30-31] 中, 用以实现模型无关的最优一致性控制。然而, 这些文献主要关注无限时域一致性控制问题。相比之下, 模型无关的有限时域一致性控制问题更具挑战性, 因为它需要在满足值函数终端约束条件的同时求解耦合的时变 HJB 方程。

为解决上述问题, 学者们开始研究基于 ADP

或 RL 的算法, 以逼近耦合的时变 HJB 方程的近似解, 从而实现 MASs 的有限时域最优一致性控制。文献 [32] 提出一种基于局部动力学的离策略 (Off-policy) RL 算法, 实现线性 MASs 的有限时域最优状态一致性控制。此外, 文献 [33] 针对非线性 MASs 提出基于 ADP 的有限时域鲁棒事件触发最优状态一致性控制方法。然而, 上述一致性控制器的设计^[32-33] 仍然依赖于 MASs 的部分模型信息, 而在实际情况下, 这些系统模型信息通常难以获得。

为克服系统模型必须已知的问题, 文献 [34] 采用神经网络逼近每个智能体的动态特性, 然后在神经网络模型的基础上基于 ADP 设计有限时域最优编队控制方法。然而, 这种方式会产生额外的计算开销, 并引入逼近误差, 从而影响 ADP 方法的有效性。文献 [35] 提出一种基于积分 RL 算法和零和博弈理论的模型无关有限时域鲁棒最优编队包含控制方法。

由于在实际的 MASs 中普遍存在执行器饱和的问题, 如无人车电机的输出转矩受最大功率限制, 无人机的舵面受物理结构限制等, 饱和的非线性特性通常会导致系统性能下降, 甚至可能导致系统不稳定, 使得执行器饱和问题在理论和实践上都极具挑战性。上述研究结果 [32-35] 无法确保在模型未知的情况下实现具有执行器饱和约束的 MASs 一致性控制。

为解决这一问题, 学者们提出基于 RL 或 ADP 的方法来处理执行器饱和的 MASs 模型无关一致性控制问题。例如, 文献 [36] 提出一种新型的辨识-评价-执行结构, 结合粘性消失法, 解决了有输入约束 MASs 的领导-跟随最优一致性控制问题。文献 [37] 提出一种离策略 RL 算法, 通过逼近求解具有非二次型代价函数的耦合 HJB 方程, 以实现一致性控制。文献 [31, 38-40] 使用低增益反馈 (Low gain feedback, LGF) 方法^[41] 处理执行器饱和问题, 并结合 ADP 方法实现执行器饱和的 MASs 最优一致性控制。然而, 这些基于 ADP 或 RL 的模型无关一致性控制方法主要解决的是存在执行器饱和的 MASs 无限时域一致性控制问题, 只能实现渐近一致性控制, 即理论调控时间趋于无穷。文献 [42] 基于 ADP 研究具有对称或不对称输入约束条件的 MASs 事件驱动有限时域最优状态一致性控制问题, 但其控制器的设计要求已知系统的模型信息。

受上述分析的启发, 本文将 LGF 方法与 Q 学习相结合, 用以解决执行器饱和的离散时间线性 MASs 模型无关有限时域一致性控制问题。首先, 根据 LGF 方法的思想, 推导得到修正的时变黎卡提方程 (Modified time-varying Riccati equation,

MTVRE). 求解 MTVRE 可以得到时变的低增益反馈律, 同时可以通过调整低增益参数来避免执行器饱和. 然后, 参考文献 [43–44], 设计依赖于系统状态、控制输入和低增益参数的时变参数化 Q 函数 (Time-varying parameterized Q-function, TVP-QF). 在 TVPQF 的基础上, 提出一种基于 Q 学习后向时间迭代模型无关一致性控制方法, 该方法在不需要已知系统动力学模型的前提下, 能够逼近求解 MTVRE, 从而实现离散时间 MASs 的有限时域一致性控制.

本文将 LGF 方法与 Q 学习相结合, 提出一种针对执行器饱和的模型无关有限时域一致性控制方法. 主要贡献如下: 设计一种依赖于智能体状态、控制输入和低增益参数的 TVPQF. 基于 TVPQF, LGF 控制器的设计减少了对系统动力学模型的依赖; 提出一种可以迭代更新低增益参数的后向时间模型无关控制算法, 并证明所提算法得到的时变 LGF 控制增益矩阵收敛于 MTVRE 的解; 另外, 证明所提算法不仅可以实现半全局一致性, 而且可以保证执行器饱和条件下的全局一致性控制, 并通过仿真实验进行论证.

本文的结构安排如下: 第 1 节首先介绍代数图论的相关知识, 并结合 LGF 方法介绍执行器饱和的离散时间 MASs 有限时域一致性控制问题的基于模型的求解方案. 第 2 节首先证明可以将执行器饱和的离散时间 MASs 有限时域一致性控制问题转化为执行器饱和的单智能体的最优控制问题, 接着提出基于 TVPQF 的后向时间迭代算法以逼近求解最优控制问题对应的 MTVRE. 第 3 节提供仿真结果验证本文方法的有效性, 并进行对比实验, 比较性能指标突显本文方法的优越性. 第 4 节为结束语.

符号说明: $\mathbf{R}$ 表示实数集, $\mathbf{R}^{n \times m}$ 表示 $n \times m$ 维矩阵. I 表示具有兼容维数的单位矩阵. $\mathbf{0}$ 表示具有兼容维数的全零向量或矩阵. $\lambda_i(A)$ 表示矩阵 A 的第 i 个特征值. Re 表示实部. $\text{rank}(A)$ 表示矩阵 A 的秩. argmax 表示最大值索引. argmin 表示最小值索引. vec 为矩阵的拉直运算, 把矩阵按照列的顺序一列接一列的组成一个长向量. x^T 表示向量 x 的转置.

1 预备知识

1.1 代数图论

有 N 个节点的加权图可记为 $G = (V, E, D)$, 其中节点和边的集合记为 $V = \{v_1, v_2, \dots, v_N\}$ 和 $E = \{(v_i, v_j) : v_i, v_j \in V\}$. 节点之间的连接关系由

行随机矩阵 $D = [d_{ij}] \in \mathbf{R}^{N \times N}$ 决定, 其中 $d_{ii} > 0$, $\sum_{j=1}^N d_{ij} = 1$. 如果 $(v_i, v_j) \in E$, $d_{ij} > 0$; 否则 $d_{ij} = 0$. 对于无向图 G , 行随机矩阵 D 是对称的, 如果在任何一对不同的节点之间存在一条路径, 则称无向图 G 是连通的. $I - D$ 可看作是一种特殊的拉普拉斯矩阵, 满足 $\text{Re}(\lambda_1(I - D)) < \text{Re}(\lambda_2(I - D)) \leq \dots \leq \text{Re}(\lambda_N(I - D))$. 此外, 当且仅当有向图 G 包含一个有向生成树, 或无向图 G 连通时, 1 是 D 的一个单特征值. 令 $r \in \mathbf{R}^N$ 表示与 $I - D$ 的零特征值相关的左特征向量, 其满足 $r^T \mathbf{1} = 1$.

1.2 问题描述

考虑由 N 个执行器饱和的智能体组成的离散时间 MASs:

$$x_i(k+1) = Ax_i(k) + B\rho(u_i(k)), \quad i = 1, 2, \dots, N \quad (1)$$

式中, $x_i(k) \in \mathbf{R}^n$, $u_i(k) \in \mathbf{R}^m$ 分别表示智能体 i 的状态向量以及输入向量; $\rho(\cdot) : \mathbf{R}^m \rightarrow \mathbf{R}^m$ 表示饱和函数, 对于 $j = 1, 2, \dots, m$ 满足:

$$\rho(u_i^j(k)) = \begin{cases} -c, & u_i^j(k) < -c \\ u_i^j(k), & -c \leq u_i^j(k) \leq c \\ c, & u_i^j(k) > c \end{cases} \quad (2)$$

式中, c 表示饱和极限.

假设 1. 本文中, 智能体的系统模型是确定且未知的, 即 $A \in \mathbf{R}^{n \times n}$, $B \in \mathbf{R}^{n \times m}$ 表示确定性的未知系统矩阵.

假设 2. 系统矩阵 (A, B) 为输入有界下渐近零可控 (Asymptotically null controllable with bounded controls, ANCBC), 即系统 (A, B) 是可控的, 且 A 的所有特征值都在单位圆上或单位圆内^[41].

假设 3. 本文所考虑的用以描述离散时间 MASs (1) 拓扑结构的无向图 G 是连通的.

假设 4. 本文所考虑的离散时间 MASs (1) 的阶次已知, 即 n 是已知的.

本文研究的是具有执行器饱和的离散时间 MASs 的有限时域一致性控制问题. 所考虑的具体问题是: 在有限的时间区间内, 通过适当的控制策略设计, 使得所有智能体的状态在终端时刻达到一致, 即 $\lim_{k \rightarrow \tau} \|x_i(k) - x_j(k)\| = 0$. 这种有限时域一致性控制要求在给定的时间范围 τ 内, 使所有智能体的状态在终端时刻达到某个共同的期望状态, 而不是在无限时域上渐近趋于一致.

参考文献 [31], 针对离散时间 MASs (1) 可以设计如下状态反馈控制律:

$$u_i(k) = K(k) \sum_{j=1}^N d_{ij} (x_i(k) - x_j(k)) \quad (3)$$

其中, $K(k)$ 为待设计的反馈控制增益矩阵.

引理 1. 对于具有 N 个节点的离散时间 MASs (1), 如果其对应的无向图 G 是联通的, 则有 $\mu = 4/(N(N-1)) \leq \lambda_2(I-D)^{[45]}$.

引理 2. 如果假设 2 和假设 3 成立, 则对于给定的有界集 $\mathcal{X} \in \mathbf{R}^n$, $\forall x_i(0) \in \mathcal{X}$, $i = 1, 2, \dots, N$, 存在最优低增益参数 $\varepsilon^* \in (0, 1]$, 对于任意 $\varepsilon \in (0, \varepsilon^*]$, 离散时间 MASs (1) 可以在控制协议 (3) 下实现半全局一致性, 其中最优反馈控制增益矩阵满足:

$$K_\varepsilon^*(k) = -(B^T P_\varepsilon^*(k+1)B + I)^{-1} B^T P_\varepsilon^*(k+1)A \quad (4)$$

式中, $P_\varepsilon^*(k)$ 满足如下 MTVRE:

$$P_\varepsilon^*(k) = A^T P_\varepsilon^*(k+1)A + \varepsilon I - (2\mu - \mu^2) \times \\ A^T P_\varepsilon^*(k+1)B(B^T P_\varepsilon^*(k+1)B + I)^{-1} \times \\ B^T P_\varepsilon^*(k+1)A \quad (5)$$

同时, $\lim_{\varepsilon \rightarrow 0} P_\varepsilon^*(k) = 0$ 是单调的^[38].

注 1. 文献 [38] 考虑的是无限时域 MASs 一致性控制问题, 需求解修正的时变黎卡提方程. 而本文考虑的是有限时域一致性控制问题, 需求解 MTVRE (5), 得到的正定矩阵 $P_\varepsilon^*(k)$ 以及 LGF 矩阵 $K_\varepsilon^*(k)$ 是时变的. 同时, 结合文献 [38] 中的引理 2 以及文献 [46], 容易推导得到引理 2.

注 2. 相比于式 (3), 式 (4) 中 $K_\varepsilon^*(k)$ 加下标 ε 的原因在于, 根据 LGF 方法的思想, 反馈控制增益矩阵 $K_\varepsilon^*(k)$ 可以通过 ε 进行调整, 从而使控制输入满足执行器饱和约束.

由引理 2 可知, 求解 MTVRE (5) 需要已知系统的模型参数 (A, B) . 然而, 在实际应用中, 系统的精确模型信息往往难以获取, 即便通过系统辨识可以获得模型信息, 但不可避免地会引入辨识误差. 同时, 引理 2 中给出的求解 MTVRE (5) 的方法只能实现半全局一致性. 为了解决这一问题, 本文首先将 MASs 的有限时域一致性控制问题转化为单智能体的有限时域最优控制问题, 并在无需系统模型信息且不依赖系统辨识的前提下, 提出一种结合低增益反馈与 Q 学习的模型无关数据驱动控制方法. 该方法能够在面对执行器饱和的情况下, 动态调整低增益参数, 从而在任意给定的智能体初始状态下, 实现离散时间 MASs (1) 的有限时域全局一致性控制.

2 数据驱动有限时域一致性控制

在本节中, 将首先介绍使用 LGF 方法求解执

行器饱和的单个智能体的优化控制问题, 进而推导出 MTVRE (5). 然后, 将介绍如何利用数据驱动方法, 通过单个智能体的可测量数据, 在系统模型信息未知的情况下, 逼近 LGF 控制增益矩阵 $K_\varepsilon^*(k)$, 从而实现离散时间执行器饱和 MASs (1) 的有限时域一致性控制.

2.1 执行器饱和的单智能体优化控制方法

考虑如下执行器饱和的离散时间系统:

$$x_i(k+1) = Ax_i(k) + B\varrho(\zeta_i(k)) \quad (6)$$

其中, $\zeta_i(k)$ 表示新的控制输入. 在接下来的基于 Q 学习的算法中, 将使用它来学习 LGF 矩阵 $K_\varepsilon(k)$.

定义如下有限时域性能指标:

$$J_i = \sum_{k=0}^{\tau-1} r_i(x_i(k), \zeta_i(k), \varepsilon) + \varepsilon x_i^T(\tau)x_i(\tau) \quad (7)$$

式中, 最后一项 $\varepsilon x_i^T(\tau)x_i(\tau)$ 代表终端约束条件; $r_i(x_i(k), \zeta_i(k), \varepsilon)$ 表示智能体 i 的效用函数:

$$r_i(x_i(k), \zeta_i(k), \varepsilon) = \varepsilon x_i^T(k)x_i(k) + \zeta_i^T(k)\zeta_i(k) \quad (8)$$

根据有限时域性能指标 (7), 每个智能体 i 的值函数可以表示为:

$$V_i(x_i(k)) = \sum_{j=k}^{\tau-1} r_i(x_i(j), \zeta_i(j), \varepsilon) + \varepsilon x_i^T(\tau)x_i(\tau) \quad (9)$$

下面引理证明, 当控制输入 $\zeta_i(k) = \mu K_\varepsilon(k) \times x_i(k)$ 时, 值函数 (9) 可以表示为二次型形式.

引理 3. 如果离散时间系统 (6) 的控制输入可以表示为 $\zeta_i(k) = \mu K_\varepsilon(k)x_i(k)$, 则智能体 i 的值函数 $V_i(x_i(k))$ 可以表示为如下二次型形式:

$$V_i(x_i(k)) = x_i^T(k)P_\varepsilon(k)x_i(k) \quad (10)$$

式中, $P_\varepsilon(k) = P_\varepsilon^T(k) > 0$. $P_\varepsilon(\tau) = \varepsilon I$.

证明. 本部分将基于最优性原理, 利用终端约束条件采用后向时间的方式进行证明.

当 $k = \tau$ 时, 可以很容易地从式 (9) 得到:

$$V_i(x_i(\tau)) = \varepsilon x_i^T(\tau)x_i(\tau) \quad (11)$$

因此, 可以得到 $P_\varepsilon(\tau) = P_\varepsilon^T(\tau) = \varepsilon I$.

当 $k = \tau - 1$ 时, 结合式 (8) 和 (9) 可以得到:

$$V_i(x_i(\tau-1)) = \varepsilon x_i^T(\tau-1)x_i(\tau-1) + \\ \zeta_i^T(\tau-1)\zeta_i(\tau-1) + \varepsilon x_i^T(\tau)x_i(\tau) \quad (12)$$

将式 (6) 代入式 (12) 中, 得到:

$$\begin{aligned}
V_i(x_i(\tau-1)) = & \varepsilon x_i^T(\tau-1)x_i(\tau-1) + \\
& \zeta_i^T(\tau-1)\zeta_i(\tau-1) + \\
& \varepsilon(Ax_i(\tau-1) + B\zeta_i(\tau-1))^T \times \\
& (Ax_i(\tau-1) + B\zeta_i(\tau-1)) \quad (13)
\end{aligned}$$

然后, 将 $\zeta_i(\tau-1) = \mu K_\varepsilon(\tau-1)x_i(\tau-1)$ 代入式 (13) 中, 可以得到:

$$\begin{aligned}
V_i(x_i(\tau-1)) = & x_i^T(\tau-1)[\varepsilon I + \mu^2 K_\varepsilon^T(\tau-1)K_\varepsilon(\tau-1) \\
& + \varepsilon(A + \mu BK_\varepsilon(\tau-1))^T \times \\
& (A + \mu BK_\varepsilon(\tau-1))]x_i(\tau-1) \quad (14)
\end{aligned}$$

当 $k = \tau-1$ 时, 从式 (14) 可以得到:

$$\begin{aligned}
P_\varepsilon(\tau-1) = & \varepsilon I + \mu^2 K_\varepsilon^T(\tau-1)K_\varepsilon(\tau-1) + \\
& \varepsilon(A + \mu BK_\varepsilon(\tau-1))^T \times \\
& (A + \mu BK_\varepsilon(\tau-1)) \quad (15)
\end{aligned}$$

从上式可以得到 $P_\varepsilon(\tau-1) = P_\varepsilon^T(\tau-1) > 0$.

采用与 $P_\varepsilon(\tau-1)$ 相同的方式, 可以类似地确定, 对于 $k = 0, 1, \dots, \tau-2$, 矩阵 $P_\varepsilon(k)$ 也符合 $P_\varepsilon(k) = P_\varepsilon^T(k) > 0$. $\square$

下面定理将证明, 针对执行器饱和的离散时间系统 (6) 以及对应的有限时域性能指标 (7), 存在最优的 LGF 控制增益矩阵 $K_\varepsilon(k)$, 使得智能体 i 的值函数 $V_i(x_i(k))$ 可以表示为式 (10).

定理 1. 考虑执行器饱和和离散时间系统 (6) 以及对应的有限时域性能指标 (7), 其最优控制律满足:

$$\zeta_i^*(k) = K_\varepsilon^*(k)x_i(k) \quad (16)$$

其中, $K_\varepsilon^*(k)$ 满足式 (4). 如果令 $\zeta_i^*(k) = \mu K_\varepsilon^*(k)x_i(k)$, 则 $P_\varepsilon^*(k)$ 满足式 (5).

证明. 根据值函数的定义 (9) 可知, 值函数满足如下贝尔曼方程:

$$V_i(x_i(k)) = \varepsilon x_i^T(k)x_i(k) + \zeta_i^T(k)\zeta_i(k) + V_i(x_i(k+1)) \quad (17)$$

同时, 最优值函数满足:

$$\begin{aligned}
V_i^*(x_i(k)) = & \min_{\zeta_i(k)} \sum_{j=k}^{\tau-1} (\varepsilon x_i^T(j)x_i(j) + \zeta_i^T(j)\zeta_i(j) + \\
& \varepsilon x_i^T(\tau)x_i(\tau)) \quad (18)
\end{aligned}$$

结合式 (17) 和式 (18), 可以得到如下的贝尔曼最优方程:

$$\begin{aligned}
V_i^*(x_i(k)) = & \min_{\zeta_i(k)} (\varepsilon x_i^T(k)x_i(k) + \zeta_i^T(k)\zeta_i(k) + \\
& V_i^*(x_i(k+1))) \quad (19)
\end{aligned}$$

当 $k = \tau-1$ 时, 由式 (18) 可知:

$$\begin{aligned}
V_i^*(x_i(\tau-1)) = & \min_{\zeta_i(\tau-1)} (\varepsilon x_i^T(\tau-1)x_i(\tau-1) + \\
& \zeta_i^T(\tau-1)\zeta_i(\tau-1) + \varepsilon x_i^T(\tau)x_i(\tau)) \quad (20)
\end{aligned}$$

将式 (6) 代入式 (20) 中, 得到:

$$\begin{aligned}
V_i^*(x_i(\tau-1)) = & \min_{\zeta_i(\tau-1)} (\varepsilon x_i^T(\tau-1)x_i(\tau-1) + \\
& \zeta_i^T(\tau-1)\zeta_i(\tau-1) + \\
& \varepsilon(Ax_i(\tau-1) + B\zeta_i(\tau-1))^T \times \\
& (Ax_i(\tau-1) + B\zeta_i(\tau-1))) \quad (21)
\end{aligned}$$

从式 (21) 可以得到最优控制策略满足:

$$\begin{aligned}
\zeta_i^*(\tau-1) = & \arg \min_{\zeta_i(\tau-1)} (\varepsilon x_i^T(\tau-1)x_i(\tau-1) + \\
& \zeta_i^T(\tau-1)\zeta_i(\tau-1) + \\
& \varepsilon(Ax_i(\tau-1) + B\zeta_i(\tau-1))^T \times \\
& (Ax_i(\tau-1) + B\zeta_i(\tau-1))) \quad (22)
\end{aligned}$$

为了得到最优控制策略, 可以通过上式右半部分对 $\zeta_i(\tau-1)$ 求导, 并令导数为零. 则有:

$$2\zeta_i^T(\tau-1) + 2\varepsilon(Ax_i(\tau-1) + B\zeta_i(\tau-1))^T B = 0 \quad (23)$$

因此, 可以得到最优控制策略:

$$\zeta_i^*(\tau-1) = -\varepsilon(\varepsilon B^T B + I)^{-1} B^T Ax_i(\tau-1) \quad (24)$$

结合值函数的终端约束条件可知 $P_\varepsilon^*(\tau) = \varepsilon I$, 则式 (24) 可以重写为:

$$\begin{aligned}
\zeta_i^*(\tau-1) = & (B^T P_\varepsilon^*(\tau)B + I)^{-1} \times \\
& B^T P_\varepsilon^*(\tau)Ax_i(\tau-1) = \\
& K_\varepsilon^*(\tau-1)x_i(\tau-1) \quad (25)
\end{aligned}$$

比较式 (4) 和式 (25), 可知 $K_\varepsilon^*(\tau-1)$ 满足式 (4).

结合文献 [46] 以及引理 3, 可以得到最优值函数 $V_i^*(x_i(\tau-1))$ 可以写成如下形式:

$$V_i^*(x_i(\tau-1)) = x_i^T(\tau-1)P_\varepsilon^*(\tau-1)x_i(\tau-1) \quad (26)$$

同时, 将 $\zeta_i^*(\tau-1) = \mu K_\varepsilon^*(\tau-1)x_i(\tau-1)$ 代入式 (20) 中, 很容易得到 $P_\varepsilon^*(\tau-1)$ 满足式 (15).

采用与 $k = \tau-1$ 相同的方式, 可以依次得到 $K_\varepsilon^*(k)$, $k = \tau-2, \dots, 1, 0$ 满足式 (4), 并且值函数满足:

$$V_i^*(x_i(k)) = x_i^T(k)P_\varepsilon^*(k)x_i(k) \quad (27)$$

此外, 将 $\zeta_i^*(k) = \mu K_\varepsilon^*(k)x_i(k)$ 代入式 (19) 中, 并结合式 (27), 很容易得到 $P_\varepsilon^*(k)$ 满足式 (15). $\square$

注 3. 与有限/固定时间控制不同, 本文所考虑的有限时域一致性控制是指控制器在一个预算的时

间段内进行设计. 在这个时间段结束时, 控制器的目标是使系统状态达到某个期望的状态或者满足特定的性能指标. 有限时域控制问题通常涉及优化一个性能指标函数, 该函数定义在从初始时刻到终止时刻的时间区间上, 如本文所考虑的有限时域性能指标函数 (7), 并且需要考虑在此期间系统的动态行为和可能存在的约束条件, 如本文所考虑的执行器饱和和约束. 而有限时间控制强调的是收敛时间 t 趋于一个固定值 T 达到稳定, 该 T 是根据初值和控制参数计算出来的. 固定时间控制是一种特殊的有限时间控制, 也是 t 趋于一个固定值 T 达到稳定, 该 T 的计算和初值无关, 但是计算的 T 有保守性. 有限时域控制可以看作有限时间控制的一种特殊情况, 其侧重点在于需要在固定时间范围内优化一个性能指标函数.

注 4. 根据低增益反馈控制方法^[41], 可以对低增益参数进行动态调整, 逐步将控制输入限制在饱和值范围内, 从而避免执行器饱和现象. 在引理 3 以及定理 1 的证明过程中, 由于低增益参数的存在, 在证明过程中假定通过调整低增益参数得到满足执行器饱和和约束的控制输入. 因此, 在涉及控制输入的证明过程中, 饱和函数 $\rho(\cdot)$ 没有显示地出现.

从以上分析可知, 可以将针对执行器饱和的离散时间 MASs (1) 的有限时域一致性控制问题转化为针对执行器饱和的离散时间系统 (6) 以及有限时域性能指标 (7) 的最优控制问题. 不同之处在于, 为了实现有限时域一致性控制, 需要改变由最优控制问题求得的控制策略. 同时, 依据 LGF 方法的特点, 可以通过调整低增益参数 ε 实现避免执行器饱和的目标.

2.2 模型无关有限时域控制方法

在这一部分, 首先, 结合 Q 学习的思想定义 TVPQF; 然后, 提出一种数据驱动的后向时间迭代方法, 在仅需要单个智能体可测量数据的前提下, 逼近求解 MTVRE (5), 以实现有限时域一致性控制.

依据文献 [27], 定义如下 TVPQF:

$$Q_\varepsilon(x_i(k), \zeta_i(k), \tau - k) = r_i(x_i(k), \zeta_i(k), \varepsilon) + V_i^*(x_i(k+1)) \quad (28)$$

其中, $Q_\varepsilon(x_i(\tau)) = \varepsilon x_i^T(\tau)x_i(\tau)$.

定义变量 $\xi_i(k) = [x_i^T(k), \zeta_i^T(k)]^T$. 同时, 将式 (6) 和 (19) 代入式 (28) 中, 可以得到:

$$Q_\varepsilon(x_i(k), \zeta_i(k), \tau - k) = \xi_i^T(k)\mathcal{H}_\varepsilon(k)\xi_i(k) \quad (29)$$

式中, $\mathcal{H}_\varepsilon(k)$ 表示 TVPQF 的核函数, 定义如下:

$$\mathcal{H}_\varepsilon(k) := \begin{bmatrix} \mathcal{H}_{xx}(k) & \mathcal{H}_{x\zeta}(k) \\ \mathcal{H}_{\zeta x}(k) & \mathcal{H}_{\zeta\zeta}(k) \end{bmatrix} = \begin{bmatrix} \varepsilon I + A^T P_\varepsilon(k+1)A & A^T P_\varepsilon(k+1)B \\ B^T P_\varepsilon(k+1)A & B^T P_\varepsilon(k+1)B + I \end{bmatrix} \quad (30)$$

同时, 通过 TVPQF 的定义 (28) 可以得到最优值函数与最优 TVPQF 的关系如下:

$$V_i^*(x_i(k)) = \min_{\zeta_i(k)} Q_\varepsilon(x_i(k), \zeta_i(k), \tau - k) = Q_\varepsilon^*(x_i(k), \zeta_i^*(k), \tau - k) \quad (31)$$

根据 TVPQF 的定义可知, 最优 LGF 控制律满足:

$$\zeta_i^*(k) = \arg \min_{\zeta_i(k)} Q_\varepsilon(x_i(k), \zeta_i(k), \tau - k) \quad (32)$$

求解 $\frac{\partial Q_\varepsilon(x_i(k), \zeta_i(k), \tau - k)}{\partial \zeta_i(k)} = 0$, 可以得到:

$$K_\varepsilon^*(k) = -\mathcal{H}_{\zeta\zeta}^{-1}(k)\mathcal{H}_{\zeta x}^*(k) \quad (33)$$

另外, 将 $\zeta_i^*(k) = \mu K_\varepsilon^*(k)x_i(k)$ 、式 (33) 代入式 (31), 同时结合式 (29), 得到:

$$P_\varepsilon^*(k) = \mathcal{H}_{xx}^*(k) - \mu K_\varepsilon^*(k)\mathcal{H}_{\zeta x}^*(k) + \mu \mathcal{H}_{x\zeta}^*(k) \times K_\varepsilon^{*,T}(k) + \mu^2 K_\varepsilon^*(k)\mathcal{H}_{\zeta\zeta}^*(k)K_\varepsilon^{*,T}(k) \quad (34)$$

根据式 (33) 和 (34) 可知, 通过设计的 TVPQF, 可以将计算 $P_\varepsilon^*(k)$ 转变为计算 $\mathcal{H}_\varepsilon^*(k)$, 以获取最优 LGF 控制增益矩阵 $K_\varepsilon^*(k)$, 并且避免对系统模型信息的依赖. 下面将介绍如何采用后向时间的方式逼近求解 $\mathcal{H}_\varepsilon^*(k)$.

假设通过 η 次实验, 收集到 η 组样本数据 $\{x_i^j(k), \zeta_i^j(k), x_i^j(k+1)\}$, 其中 $j = 1, 2, \dots, \eta$.

当 $k = \tau - 1$ 时, 定义:

$$Q_\varepsilon^j(\tau - 1) = \varepsilon x_i^{j,T}(\tau - 1)x_i^j(\tau - 1) + \zeta_i^{j,T}(\tau - 1) \times \zeta_i^j(\tau - 1) + x_i^{j,T}(\tau)P_\varepsilon^*(\tau)x_i^j(\tau) \quad (35)$$

式中, $P_\varepsilon^*(\tau) = \varepsilon I$.

同时, 根据式 (29), 可以得到 $Q_\varepsilon^j(\tau - 1)$ 的另一种表达形式如下:

$$\hat{Q}_\varepsilon^j(\tau - 1) = \xi_i^{j,T}(\tau - 1)\mathcal{H}_\varepsilon(\tau - 1)\xi_i^j(\tau - 1) \quad (36)$$

应用线性参数化方法, 式 (36) 可以重写为:

$$\hat{Q}_\varepsilon^j(\tau - 1) = \bar{\xi}_i^{j,T}(\tau - 1)\text{vec}(\mathcal{H}_\varepsilon(\tau - 1)) \quad (37)$$

其中,

$$\begin{aligned} \bar{\xi}_i^j(\tau - 1) = & [(\xi_i^{1,j})^2, 2\xi_i^{1,j}\xi_i^{2,j}, \dots, 2\xi_i^{1,j}\xi_i^{l,j}, \\ & (\xi_i^{2,j})^2, 2\xi_i^{2,j}\xi_i^{3,j}, \dots, \\ & 2\xi_i^{2,j}\xi_i^{l,j}, \dots, (\xi_i^{l,j})^2]^T \end{aligned}$$

上面变量的表达式中 $l = n + m$ 表示变量 $\bar{\xi}_i^j(\tau -$

1) 的维数. 另外, 为方便, 省去了 $\tau - 1$.

结合式 (35) 和 (37) 可知, 可以通过求解如下优化方程用以获取 TVPQF 对应的最优核矩阵 $\mathcal{H}_\varepsilon^*(\tau - 1)$:

$$\begin{aligned} \text{vec}(\mathcal{H}_\varepsilon^*(\tau - 1)) = \arg \min \sum_{j=1}^{\eta} (\bar{\xi}_i^j, {}^T(\tau - 1) \times \\ \text{vec}(\mathcal{H}_\varepsilon(\tau - 1)) - \mathcal{Q}_\varepsilon^j(\tau - 1))^2 \end{aligned} \quad (38)$$

得到 $\mathcal{H}_\varepsilon^*(\tau - 1)$, 就可以通过式 (33) 求解最优 LGF 控制增益矩阵 $K_\varepsilon^*(\tau - 1)$, 以及通过式 (34) 获取最优值函数对应的核矩阵 $P_\varepsilon^*(\tau - 1)$.

依据求解 $\mathcal{H}_\varepsilon^*(\tau - 1)$ 的思路, 可以通过后向时间求解的方式逼近求解 $\mathcal{H}_\varepsilon^*(k)$, $K_\varepsilon^*(k)$, 以及 $P_\varepsilon^*(k)$, $k = \tau - 2, \dots, 1, 0$.

当 $k = \tau - 2, \dots, 1, 0$ 时, 定义:

$$\begin{aligned} \mathcal{Q}_\varepsilon^j(k) = \varepsilon x_i^j, {}^T(k) x_i^j(k) + \zeta_i^j, {}^T(k) \zeta_i^j(k) + \\ x_i^j, {}^T(k+1) P_\varepsilon^*(k+1) x_i^j(k+1) \end{aligned} \quad (39)$$

同样地, 可以得到 $\mathcal{Q}_\varepsilon^j(k)$ 的另一种表达形式:

$$\hat{\mathcal{Q}}_\varepsilon^j(k) = \xi_i^j, {}^T(k) \mathcal{H}_\varepsilon(k) \xi_i^j(k) \quad (40)$$

参照式 (38), 可以得到如下优化问题:

$$\begin{aligned} \text{vec}(\mathcal{H}_\varepsilon^*(k)) = \\ \arg \min \sum_{j=1}^{\eta} \left(\bar{\xi}_i^j, {}^T(k) \text{vec}(\mathcal{H}_\varepsilon(k)) - \mathcal{Q}_\varepsilon^j(k) \right)^2 \end{aligned} \quad (41)$$

通过式 (41) 求解得到 $\mathcal{H}_\varepsilon^*(k)$, 就可以通过式 (33) 求解最优 LGF 控制增益矩阵 $K_\varepsilon^*(k)$, 以及通过式 (34) 获取最优值函数对应的核矩阵 $P_\varepsilon^*(k)$. 下面将介绍如何求解优化问题 (38) 和问题 (41). 由于两者具有相似性, 下面将问题 (38) 和问题 (41) 归结为一类问题进行介绍.

优化问题 (38) 和问题 (41) 可以写成如下形式:

$$\begin{aligned} \text{vec}(\mathcal{H}_\varepsilon^*(k)) = \\ \arg \min \sum \left(\bar{\xi}_i^T(k) \text{vec}(\mathcal{H}_\varepsilon(k)) - \mathcal{Q}_\varepsilon(k) \right)^2 \end{aligned} \quad (42)$$

式中, $\bar{\xi}_i(k) = [\bar{\xi}_i^1(k), \bar{\xi}_i^2(k), \dots, \bar{\xi}_i^\eta(k)]^T$; $\mathcal{Q}_\varepsilon(k) = [\mathcal{Q}_\varepsilon^1(k), \mathcal{Q}_\varepsilon^2(k), \dots, \mathcal{Q}_\varepsilon^\eta(k)]^T$, $k = 0, 1, \dots, \tau - 1$.

应用最小二乘法, 可以得到优化问题 (42) 的解如下:

$$\text{vec}(\mathcal{H}_\varepsilon^*(k)) = (\bar{\xi}_i(k) \bar{\xi}_i^T(k))^{-1} \bar{\xi}_i(k) \mathcal{Q}_\varepsilon(k) \quad (43)$$

为确保优化问题 (42) 的解 (43) 的唯一性, 需要满足如下条件:

$$\text{rank}(\bar{\xi}_i(k)) = \frac{l(l+1)}{2} \quad (44)$$

即矩阵 $\bar{\xi}_i(k)$ 满秩.

如果搜集到的样本 $\{x_i^j(k), \zeta_i^j(k), x_i^j(k+1)\}$ 的数量 $\eta \geq l(l+1)/2$, 且每次实验收集到的数据之间服从高斯分布, 那么条件 (44) 成立^[46].

由以上分析可知, 采用后向时间求解的方式可以得到最优 TVPQF 对应的核矩阵 $\mathcal{H}_\varepsilon^*(k)$. 同时, 由式 (35) 和 (39) 可知, TVPQF 会受到低增益参数 ε 的影响. 因此, 可以通过调整低增益参数 ε 用以更新 LGF 控制增益矩阵 $K_\varepsilon^*(k)$, 从而使控制器 $u_i(k)$ 避免输入饱和. 算法 1 对上面的论述进行了总结.

算法 1. 执行器饱和和约束下模型无关有限时域一致性控制

输入. 实验次数 η , 低增益参数 ε , 有限时域 τ .

输出. 最优 LGF 控制增益矩阵 $K_\varepsilon^*(k)$, 以及最优 TVPQF 对应的核矩阵 $\mathcal{H}_\varepsilon^*(k)$, $k = 0, 1, \dots, \tau - 1$.

- 1) 数据收集: 生成符合高斯分布的随机控制输入 $\{\zeta_i^j(0), \zeta_i^j(1), \dots, \zeta_i^j(\tau - 1)\}$, 以及随机初始状态变量 $x_i^j(0)$, $j = 1, 2, \dots, \eta$, 应用于系统 (6), 从而收集产生的样本数据 $\{x_i^j(k), \zeta_i^j(k), x_i^j(k+1)\}$, 其中 $j = 1, 2, \dots, \eta$; $k = 0, 1, \dots, \tau - 1$.
- 2) 计算 $K_\varepsilon^*(\tau - 1)$: 通过式 (43) 求解优化问题 (38), 得到 $\mathcal{H}_\varepsilon^*(\tau - 1)$. 结合式 (33) 推导得到最优 LGF 控制增益矩阵 $K_\varepsilon^*(\tau - 1)$, 并将其存储.
- 3) 计算 $K_\varepsilon^*(k)$: 从 $k = \tau - 2$ 到 $k = 0$, 依次通过式 (43) 求解优化问题 (42), 迭代计算 $\mathcal{H}_\varepsilon^*(k)$. 结合式 (33) 推导得到最优 LGF 控制增益矩阵 $K_\varepsilon^*(k)$, $k = \tau - 2, \dots, 1, 0$, 并将其存储.
- 4) 饱和度检查: 对于每一个 $k = 0, 1, \dots, \tau - 1$, 验证

$$\|u_i(k)\|_\infty = \left\| K_\varepsilon^*(k) \sum_{j=1}^N d_{ij} (x_i(k) - x_j(k)) \right\|_\infty \leq c$$

其中, $i = 1, 2, \dots, N$. 如果不满足, 则减小 ε 并重复步骤 2) 和步骤 3).

- 5) 停止迭代: 当控制输入不再饱和时, 停止迭代过程.

注 5. 算法 1 中, 低增益参数 ε 可以通过比例规则进行调整: $\varepsilon_{j+1} = \alpha \varepsilon_j$, 其中 $0 < \alpha < 1$. 另外需要强调的是, 控制输入的饱和度评估发生在其应用到 MASs (1) 之前. 因此, MASs 在实际执行的过程中不会超过其执行器饱和约束.

注 6. 算法 1 中的饱和度检查环节必然会受到智能体初始状态的影响, 不同的初始状态可能会最终得到不同的低增益参数 ε . 另外, 算法 1 的目的并不是寻找最优低增益参数 ε^* , 而是对于不同的初始

状态寻找 $\varepsilon \in (0, \varepsilon^*]$, 从而得到对应的最优 LGF 控制增益矩阵 $K_\varepsilon^*(k)$, 达到避免执行器饱和的目标.

下面定理将证明通过算法 1 得到的最优 LGF 控制增益矩阵是最优的.

定理 2. 如果进行收集样本数据的实验次数 $\eta \geq l(l+1)/2$, 且收集得到的样本数据 $\{x_i^j(k), \zeta_i^j(k), x_i^j(k+1)\}$ 服从高斯分布, 则算法 1 得到的 LGF 控制增益矩阵 $K_\varepsilon^*(k)$ 是最优的, 也就是 MTVRE (5) 对应的解.

证明. 根据 LGF 方法的思想, 针对执行器饱和和约束问题, 存在最优低增益参数 ε^* [42]. 同时, 注意到算法 1 中关于低增益参数 ε 的调整处于估计 $\mathcal{H}_\varepsilon(k)$ 的外循环. 因此, 低增益参数 ε 不会影响 TVPQF 核矩阵 $\mathcal{H}_\varepsilon(k)$ 的收敛性. 假设低增益参数 ε 在算法 1 中是固定的, 即考虑 MTVRE (5) 和 TVPQF (23) 中包含相同的低增益参数 ε 的情况.

当初始样本数据 $x_i^j(0), j = 1, 2, \dots, \eta$, 以及 $\zeta_i^j(k), k = 0, 1, \dots, \tau-1$ 服从高斯分布时, 很容易得到每次收集样本数据的实验是线性独立的. 此外, 如果实验次数 $\eta \geq l(l+1)/2$, 则式 (43) 中构造得到的数据矩阵 $\bar{\xi}_i(k), k = 0, 1, \dots, \tau-1$ 满秩. 需要注意的是, 结合 $\text{vec}(\mathcal{H}_\varepsilon^*(k))$ 的定义以及式 (30) 可知, $\text{vec}(\mathcal{H}_\varepsilon^*(k)), k = 0, 1, \dots, \tau-1$ 拥有 $l(l+1)/2$ 个独立元素. 结合矩阵 $\bar{\xi}_i(k)$ 满秩的结论, 可知优化问题 (42) 有唯一解, 即为式 (43). 值得注意的是, 所设计的算法 1 以离线后向时间迭代的方式运行, 即利用终端约束条件 $P_\varepsilon(\tau)$ 从 $k = \tau-1$ 开始依次向后计算 $\text{vec}(\mathcal{H}_\varepsilon^*(k))$. 同时, 式 (43) 构成了优化问题 (42) 的唯一解. 可以得出结论: 通过执行算法 1 得到的 $\text{vec}(\mathcal{H}_\varepsilon^*(k))$ 是最优的.

值得注意的是, 矩阵 $\mathcal{H}_\varepsilon^*(k)$ 是由 $l(l+1)/2$ 个元素组成的对称矩阵, $\text{vec}(\mathcal{H}_\varepsilon^*(k))$ 表示矩阵 $\mathcal{H}_\varepsilon^*(k)$ 经过列排列之后组成的长向量. 由于算法 1 得到的 $\text{vec}(\mathcal{H}_\varepsilon^*(k))$ 是最优的. 因此, 算法 1 得到的结果 $\mathcal{H}_\varepsilon^*(k)$ 即为所定义的 TVPQF 的最优核矩阵. 结合式 (33) 以及引理 2 可知, 算法 1 得出的 LGF 控制增益矩阵 $K_\varepsilon^*(k)$ 也是最优的. 同时, 结合定理 1 以及式 (31) 可知, 通过算法 1 得到 LGF 控制增益矩阵 $K_\varepsilon^*(k)$ 等价于求解 MTVRE (5). $\square$

下面定理将证明算法 1 可以实现离散时间 MASs (1) 的全局有限时域一致性控制而不仅仅是半全局有限时域一致性控制.

定理 3. 如果假设 2 和假设 3 成立, 通过算法 1 得到的 LGF 控制增益矩阵 $K_\varepsilon^*(k)$, 离散时间 MASs (1) 可以实现全局有限时域一致性控制.

证明. 如果假设 2 和假设 3 成立, 由引理 2 以

及定理 2 可知, 算法 1 得到的 LGF 控制增益矩阵 $K_\varepsilon^*(k)$ 可以实现半全局有限时域一致性控制. 在算法 1 中, 如果控制输入违反执行器饱和, 则在下次迭代时会减小低增益参数 ε , 因此必然可以找到一个足够小的 $\varepsilon \in (0, \varepsilon^*]$ 满足执行器饱和. 另外, 从定理 2 的证明过程可知, 如果 ε 固定, 由算法 1 得到的 TVPQF 核矩阵以及 LGF 控制增益矩阵均是最优的, 且可以实现有限时域最优一致性控制. 如果智能体的初始状态不同, 必然会迭代得到一个固定的低增益参数 ε , 对应地, 即可通过算法 1 得到 LGF 控制增益矩阵. 因此, 算法 1 可以实现离散时间 MASs (1) 的全局有限时域一致性控制. $\square$

3 仿真实验

本节首先建立一个仿真实验, 来说明本文方法的有效性; 然后进行对比实验, 用本文方法与对比方法进行仿真实验, 用评价指标结果说明本文方法的优越性.

3.1 仿真实验 1

考虑一个由 6 个智能体组成的离散时间 MASs, 其动力学方程为 (1), 相关的矩阵为:

$$A = \begin{bmatrix} 0 & 1 \\ -1 & 1 \end{bmatrix}, \quad B = \begin{bmatrix} -1 \\ 0 \end{bmatrix} \quad (45)$$

矩阵 A 的特征值 $0.5 \pm 0.866i$ 都在单位圆内, 且 (A, B) 是可控的. 因此, 假设 2 成立. 在本节仿真中, 执行器饱和函数的饱和阈值设为 $c = 1$. 离散时间 MASs 的通信拓扑可以用图 1 所示的无向图表示. 从图中可以得到, 所对应的无向图是连通的. 因此, 假设 3 成立.

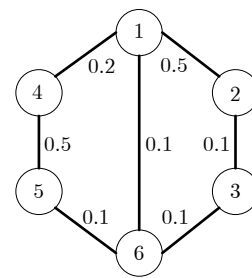


图 1 仿真 1 中 MASs 的通信拓扑

Fig. 1 MASs communication topology in simulation 1

下面将使用三个实例来说明本文所提方法的有效性. 在所有的三个实例中, 算法 1 中具体的参数设置如下: 收集样本数据的实验次数 $\eta = 100 > (3 \times 4)/2 = 6$, 初始低增益参数 $\varepsilon = 1$. 同时, 使用注 5 中的方法对 ε 进行更新, 选择 $\alpha = 0.5$. 后续将通过改

变不同的初始状态来说明算法 1 的有效性.

例 1. 在本例中, 将所有智能体的初始状态设置为 $[-1, 1] \times [-1, 1]$, 有限时域设置为 $\tau = 100$, 然后将算法 1 应用于 MASs (45) 中, 最终得到低增益参数 $\varepsilon = 0.5$. 同时, 将对应的最优 LGF 控制增益矩阵 $K_{\varepsilon}^*(k)$ 应用于系统中, 得到的 6 个智能体的系统状态如图 2 所示, 系统控制输入如图 3 所示. 从图 2 和图 3 可知, 通过算法 1 得到的控制输入可以实现有限时域一致性控制, 并且避免输入饱和.

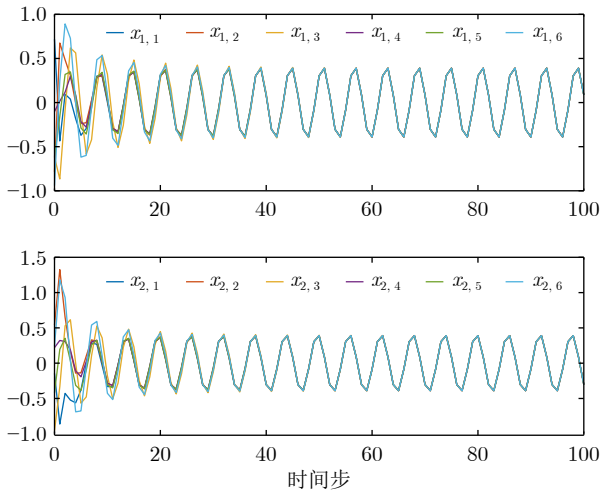


图 2 例 1 中智能体的状态

Fig.2 The states of agents in example 1

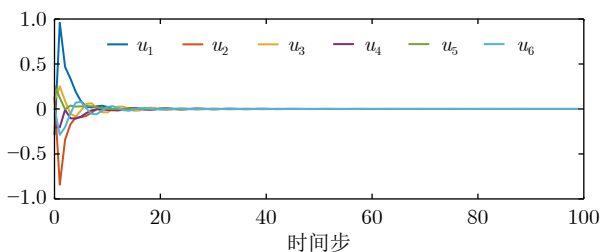


图 3 例 1 中智能体的控制输入

Fig.3 The control inputs of agents in example 1

例 2. 在本例中, 将所有智能体的初始状态设置为 $[-10, 10] \times [-10, 10]$, 有限时域设置为 $\tau = 300$, 然后将算法 1 应用于 MASs (45) 中, 最终得到低增益参数 $\varepsilon = 0.002$. 同时, 将对应的最优 LGF 控制增益矩阵 $K_{\varepsilon}^*(k)$ 应用于系统中, 得到的 6 个智能体的系统状态如图 4 所示, 系统控制输入如图 5 所示. 不同于例 1, 例 2 中智能体的初始状态的范围变大, 必然会影响 MASs 的一致性控制效果. 相比而言, 例 2 中智能体实现一致性控制的时间更长, 得到的低增益参数更小. 然而, 从图 4 和图 5 可知, 通过算法 1 得到的控制输入仍然可以实现有限时域一致性

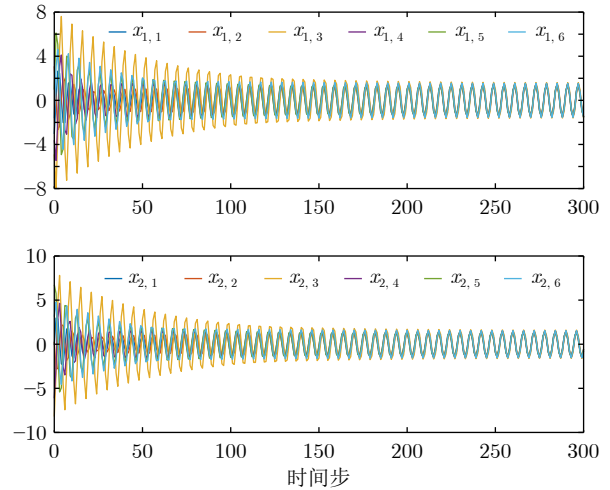


图 4 例 2 中智能体的状态

Fig.4 The states of agents in example 2

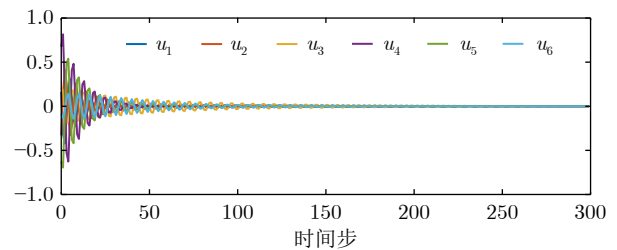


图 5 例 2 中智能体的控制输入

Fig.5 The control inputs of agents in example 2

控制, 并避免输入饱和.

例 3. 在本例中, 进一步加大了智能体初始状态的范围, 设置为 $[-100, 100] \times [-100, 100]$, 有限时域设置为 $\tau = 1\,500$, 然后将算法 1 应用于 MASs (45) 中, 最终得到低增益参数 $\varepsilon = 1.220\,7 \times 10^{-4}$. 同时, 将对应的最优 LGF 控制增益矩阵 $K_{\varepsilon}^*(k)$ 应用于系统中, 得到的 6 个智能体的系统状态如图 6 所示, 系统控制输入如图 7 所示. 从所得结果可知, 所提方法可以在有限时域内实现一致性控制, 并避免输入饱和.

以上三个例子证明了本文所提算法的有效性, 同时说明了如果智能体的初始状态越大, 控制输入需要配合越小的 LGF 控制增益矩阵 $K_{\varepsilon}(k)$ 以避免输入饱和, 因此低增益参数 ε 将会迭代更多的次数, 从而得到更小的输入值. 此外, 在输入饱和度相等的情况下 ($c = 1$), 初始状态越大, 智能体实现一致性的速度越慢, 如图 2、图 4 和图 6 所示. 通过以上三个例子, 也对定理 3 进行了验证.

3.2 仿真实验 2

在本节将所提模型无约束有限时域一致性控制算

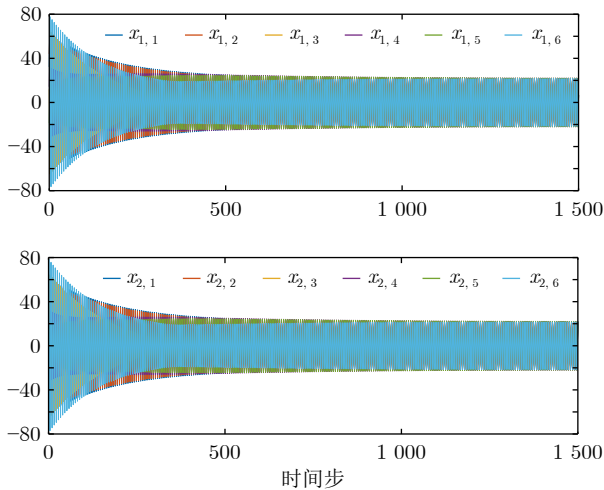


图 6 例 3 中智能体的状态

Fig.6 The states of agents in example 3

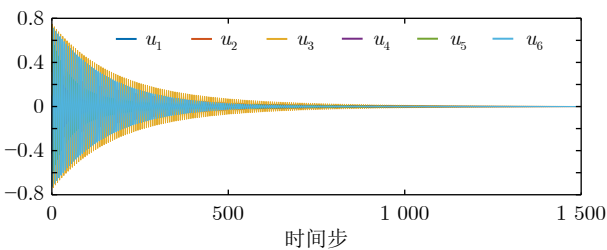


图 7 例 3 中智能体的控制输入

Fig.7 The control inputs of agents in example 3

法与文献 [38] 针对执行器饱和的模型无关无限时域一致性控制方法进行对比.

考虑一个由 5 个智能体组成的离散时间 MASs, 其动力学方程为 (1), 相关的矩阵为:

$$A = \begin{bmatrix} 0.995 & -0.194 \\ 0.194 & 0.995 \end{bmatrix}, \quad B = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \quad (46)$$

矩阵 A 的特征值 $0.9801 \pm 0.1987i$ 都在单位圆上, 且 (A, B) 是可控的. 因此, 假设 2 成立. 在本节仿真中, 执行器饱和函数的饱和阈值设为 $c = 1$. 离散时间 MASs 的通信拓扑用图 8 所示的无向图表示. 从图 8 中可以得到, 所对应的无向图是连通的. 因此, 假设 3 成立.

针对本文所提算法 1 的相关参数设置如下: 有限时域 $\tau = 120$, 收集样本数据的实验次数 $\eta = 100$, 初始低增益参数 $\varepsilon = 1$, 低增益参数 ε 调节参数 $\alpha = 0.9$. 参考文献 [38] 所提无限时域算法的相关参数设置, 初始低增益参数 $\varepsilon = 1$, $M^0 = I$, $K^0 = [0, 0]$, 收集样本数据数量 $H = 100$, 算法收敛参数设置为 0.000 01. 低增益参数 ε 的更新规则和本文所提算法 1 一致.

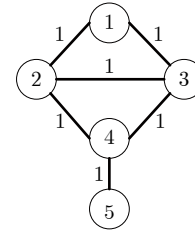


图 8 仿真 2 中 MASs 的通信拓扑

Fig.8 MASs communication topology in simulation 2

例 1. 在本例中, 首先设定 5 个智能体的初始状态为 $x_1(0) = [2.5, -2.5]^T$, $x_2(0) = [-1.5, 2]^T$, $x_3(0) = [-2, -3]^T$, $x_4(0) = [-2, -2]^T$, $x_5(0) = [1.5, 1.5]^T$. 两种算法得到的最终低增益参数均为 $\varepsilon = 3.4 \times 10^{-3}$. 采用文献 [38] 中所提算法得到的最优 LGF 控制增益矩阵 $K_\varepsilon^* = [-0.0937, -0.0730]^T$. 将两种算法得到的最优 LGF 控制增益矩阵 $K_\varepsilon^*(k)$ 和 K_ε^* 应用于 MASs (23) 中. 为了对比两种算法的一致性控制效果, 引入一致性控制误差 $\varepsilon_i(k) = \sum_{j=1}^N d_{ij}(x_i(k) - x_j(k))$. 仿真结果见图 9 和图 10.

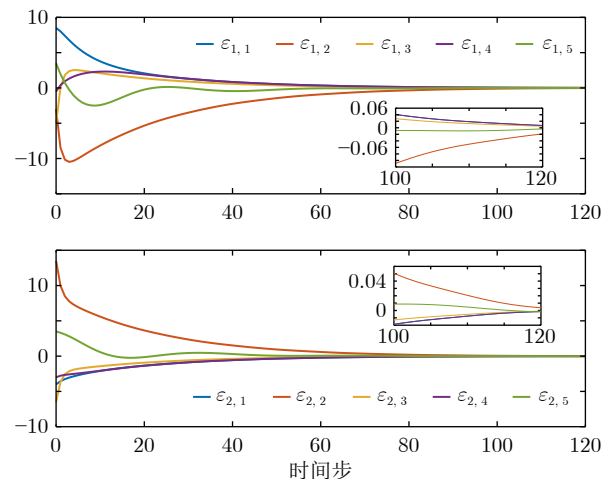


图 9 例 1 中有限时域方法获得的一致性误差

Fig.9 Consensus errors obtained by finite-horizon method in example 1

例 2. 在本例中, 改变 5 个智能体的初始状态为 $x_1(0) = [1, 2]^T$, $x_2(0) = [-0.5, -0.1]^T$, $x_3(0) = [0.3, 2]^T$, $x_4(0) = [0.8, 0.2]^T$, $x_5(0) = [-3, -2]^T$. 两种算法得到的最终低增益参数均为 $\varepsilon = 7.1 \times 10^{-3}$. 文献 [38] 所提算法得到的最优 LGF 控制增益矩阵为 $K_\varepsilon = [-0.1324, -0.1106]^T$. 最终得到的仿真结果见图 11 和图 12.

另外, 本文用每个智能体对应一致性误差的绝对误差积分 (Integral absolute error, IAE) 的平均值和均方误差 (Mean square error, MSE) 的和两

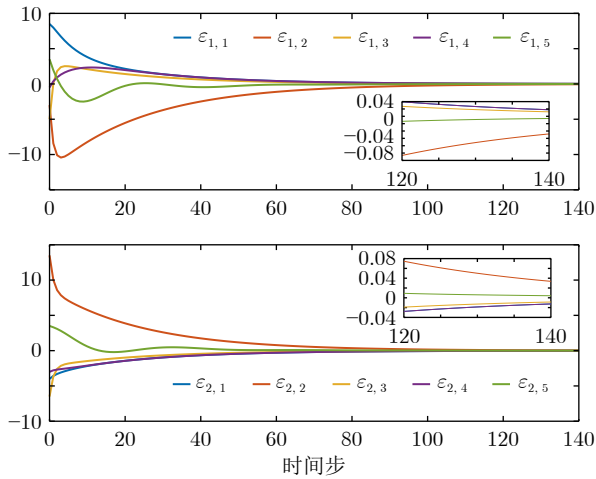


图 10 例 1 中无限时域方法获得的一致性误差

Fig.10 Consensus errors obtained by infinite-horizon method in example 1

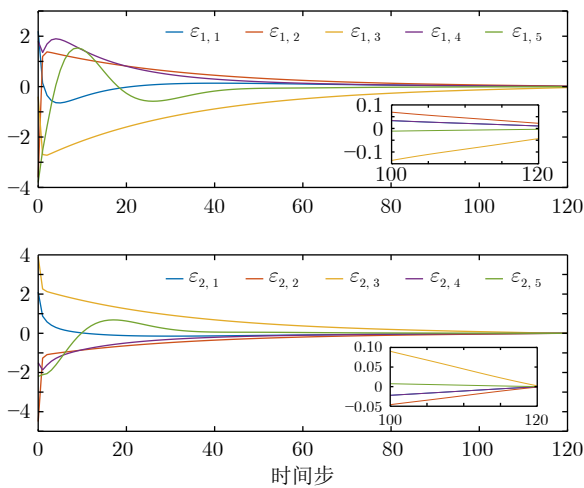


图 11 例 2 中有限时域方法获得的一致性误差

Fig.11 Consensus errors obtained by finite-horizon method in example 2

个指标^[47-48]来评价本仿真实验的控制效果, 结果见表 1.

$$\text{IAE} = \frac{\sum_{i=1}^N \sum_{k=0}^{k^*} |\varepsilon_i(k)|}{N}$$

$$\text{MSE} = \sum_{i=1}^N \frac{1}{k^*} \sum_{k=0}^{k^*} |\varepsilon_i(k)|^2$$

同时, 为了对比两种算法的一致性控制效果, 统计了智能体一致性误差对应的调节时间指标 (以一致性误差范围的 $\pm 2\%$ 进行计算), 在不同初始状态下, 将时域参数均设置为 200, 每个智能体对应的调节时间如表 2 和表 3 所示.

由图 9 ~ 图 12 以及表 1 可知, 本文所提算法

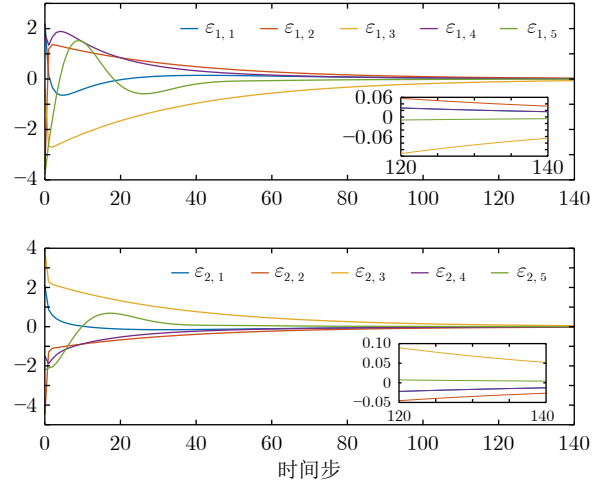


图 12 例 2 中无限时域方法获得的一致性误差

Fig.12 Consensus errors obtained by infinite-horizon method in example 2

表 1 对比实验评价指标

Table 1 Evaluation indices of comparison experiments

$100 \leq k \leq 120$	IAE	MSE
例 1-有限时域方法	0.637 7	0.005 4
例 1-无限时域方法	10.264 9	2.116 9
例 2-有限时域方法	1.074 8	0.014 7
例 2-无限时域方法	5.186 9	0.510 9

表 2 例 1 中一致性误差调节时间

Table 2 Consensus error setting time in example 1

例 1-调节时间	有限时域方法	无限时域方法
智能体 1	109	137
智能体 2	119	161
智能体 3	104	127
智能体 4	109	137
智能体 5	90	110

表 3 例 2 中一致性误差调节时间

Table 3 Consensus error setting time in example 2

例 2-调节时间	有限时域方法	无限时域方法
智能体 1	108	131
智能体 2	116	158
智能体 3	120	183
智能体 4	108	131
智能体 5	84	93

能够更快地实现一致性控制, 一致性误差较小. 同时由表 2 和表 3 可知, 在一定的时间范围内, 本文

所提的有限时域一致性控制算法得到的一致性性能指标较文献 [38] 所提无限时域一致性控制算法要好, 这也说明了本文提出算法的优越性.

4 结束语

本文提出一种基于 Q 学习的数据驱动算法, 用于求解具有未知模型参数、执行器饱和的离散时间 MASs 的有限时域一致性控制问题. 首先结合 LGF 方法, 将执行器饱和的有限时域一致性控制问题转化为执行器饱和的单智能体最优控制问题, 给出原问题的控制器设计方案. 然后在未知系统模型参数的条件下, 设计基于 Q 学习的数据驱动后向时间算法逼近求解 MTVRE, 用以获取 LGF 控制增益矩阵, 并给出该算法的收敛性说明. 最后, 给出仿真结果来验证基于 Q 学习的有限时域一致性控制算法的有效性, 并证明智能体的初始状态会影响收敛速度的问题. 同时, 还给出对比实验来评价有限时域一致性控制算法与无限时域一致性控制算法的控制效果.

在本文提出的方法中, 有限时域参数 τ 作为算法 1 的输入参数, 其在参数选择过程中需凭借经验来进行设定. 在未来的研究中, 将探讨更为精确的有限时域参数设置方法, 以确定 τ 的边界条件, 从而设定合理的有限时域参数 τ .

References

- Huang Y, Fang W T, Chen Z Y, Li Y G, Yang C H. Flocking of multiagent systems with nonuniform and nonconvex input constraints. *IEEE Transactions on Automatic Control*, 2023, **68**(7): 4329–4335
- Okine A A, Adam N, Naeem F, Kaddoum G. Multi-agent deep reinforcement learning for packet routing in tactical mobile sensor networks. *IEEE Transactions on Network and Service Management*, 2024, **21**(2): 2155–2169
- Mu C X, Liu Z Y, Yan J, Jia H J, Zhang X Y. Graph multi-agent reinforcement learning for inverter-based active voltage control. *IEEE Transactions on Smart Grid*, 2024, **15**(2): 1399–1409
- Zhao Y W, Niu B, Zong G D, Zhao X D, Alharbi K H. Neural network-based adaptive optimal containment control for non-affine nonlinear multi-agent systems within an identifier-actor-critic framework. *Journal of the Franklin Institute*, 2023, **360**(12): 8118–8143
- Fan S J, Wang T, Qin C H, Qiu J B, Li M. Optimized backstepping attitude containment control for multiple spacecrafts. *IEEE Transactions on Fuzzy Systems*, 2024, **32**(9): 5248–5258
- An L W, Yang G H, Deng C, Wen C Y. Event-triggered reference governors for collisions-free leader-following coordination under unreliable communication topologies. *IEEE Transactions on Automatic Control*, 2024, **69**(4): 2116–2130
- Wang W, Chen X. Model-free optimal containment control of multi-agent systems based on actor-critic framework. *Neurocomputing*, 2018, **314**: 242–250
- Wang W, Chen X, Fu H, Wu M. Model-free distributed consensus control based on actor-critic framework for discrete-time nonlinear multiagent systems. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 2020, **50**(11): 4123–4134
- Su H S, Miao S X. Consensus on directed matrix-weighted networks. *IEEE Transactions on Automatic Control*, 2023, **68**(4): 2529–2535
- Yang Hong-Yong, Guo Lei, Zhang Yu-Ling, Yao Xiu-Ming. Delay consensus of fractional-order multi-agent systems with sampling delays. *Acta Automatica Sinica*, 2014, **40**(9): 2022–2028 (杨洪勇, 郭雷, 张玉玲, 姚秀明. 离散时间分数阶多自主体系统的时延一致性. *自动化学报*, 2014, **40**(9): 2022–2028)
- Ma Yu-Wen, Li Xian-Wei, Li Shao-Yuan. A reduced-order protocol for linear multi-agent consensus without inter-controller communication. *Acta Automatica Sinica*, 2023, **49**(9): 1836–1844 (马煜文, 李贤伟, 李少远. 无控制器间通信的线性多智能体一致性的降阶协议. *自动化学报*, 2023, **49**(9): 1836–1844)
- He W P, Chen X, Zhang M L, Sun Y P, Sekiguchi A, She J H. Data-driven optimal consensus control for switching multiagent systems via joint communication graph. *IEEE Transactions on Industrial Informatics*, 2024, **20**(4): 5959–5968
- Zhang H P, Yue D, Dou C X, Zhao W, Xie X P. Data-driven distributed optimal consensus control for unknown multiagent systems with input-delay. *IEEE Transactions on Cybernetics*, 2019, **49**(6): 2095–2105
- Ji J W, Zhang Z C, Wang Y J, Zuo Z Q. Event-triggered consensus of discrete-time double-integrator multi-agent systems with asymmetric input saturation. *Nonlinear Dynamics*, 2024, **112**(15): 13321–13334
- Liu C, Liu L, Cao J D, Abdel-Aty M. Intermittent event-triggered optimal leader-following consensus for nonlinear multi-agent systems via actor-critic algorithm. *IEEE Transactions on Neural Networks and Learning Systems*, 2023, **34**(8): 3992–4006
- Wang J, Zhang Z T, Tian B L, Zong Q. Event-based robust optimal consensus control for nonlinear multiagent system with local adaptive dynamic programming. *IEEE Transactions on Neural Networks and Learning Systems*, 2024, **35**(1): 1073–1086
- Zattoni E. Structural invariant subspaces of singular Hamiltonian systems and nonrecursive solutions of finite-horizon optimal control problems. *IEEE Transactions on Automatic Control*, 2008, **53**(5): 1279–1284
- Ferrari-Trecate G, Galbusera L, Marciandi M, Scattolini R. A model predictive control scheme for consensus in multi-agent systems with single-integrator dynamics and input constraints. In: *Proceedings of the 46th IEEE Conference on Decision and Control*. New Orleans, USA: IEEE, 2007. 1492–1497
- Aditya P, Werner H. A distributed linear-quadratic discrete-time game approach to multi-agent consensus. In: *Proceedings of the 61st IEEE Conference on Decision and Control*. Cancun, Mexico: IEEE, 2022. 6169–6174
- Han F, Wei G L, Ding D D, Song Y. Finite-horizon H_∞ -consensus control for multi-agent systems with random parameters: The local condition case. *Journal of the Franklin Institute*, 2017, **354**(14): 6078–6097
- Li J J, Wei G L, Ding D R. Finite-horizon H_∞ consensus control for multi-agent systems under energy constraint. *Journal of the Franklin Institute*, 2019, **356**(6): 3762–3780
- Chen W, Ding D R, Dong H L, Wei G L, Ge X H. Finite-horizon H_∞ bipartite consensus control of cooperation-competition multiagent systems with round-robin protocols. *IEEE Transactions on Cybernetics*, 2021, **51**(7): 3699–3709
- Li X M, Yao D Y, Li P S, Meng W, Li H Y, Lu R Q. Secure fi-

- nite-horizon consensus control of multiagent systems against cyber attacks. *IEEE Transactions on Cybernetics*, 2022, **52**(9): 9230–9239
- 24 Powell W B. *Approximate Dynamic Programming: Solving the Curses of Dimensionality*. Hoboken: John Wiley & Sons, 2007.
- 25 Sutton R S, Barto A G. *Reinforcement Learning: An Introduction*. Cambridge: The MIT Press, 2018.
- 26 Pang Wen-Yan, Fan Jia-Lu, Jiang Yi, Lewis Frank Leroy. Optimal output regulation of partially linear discrete-time systems using reinforcement learning. *Acta Automatica Sinica*, 2022, **48**(9): 2242–2253
(庞文砚, 范家璐, 姜艺, Lewis Frank Leroy. 基于强化学习的部分线性离散时间系统的最优输出调节. *自动化学报*, 2022, **48**(9): 2242–2253)
- 27 Watkins C, Dayan P. Q-learning. *Machine Learning*, 1992, **8**(3): 279–292
- 28 Mu C X, Zhao Q, Gao Z K, Sun C Y. Q-learning solution for optimal consensus control of discrete-time multiagent systems using reinforcement learning. *Journal of the Franklin Institute*, 2019, **356**(13): 6946–6967
- 29 Liu J L, Dong Y H, Gu Z, Xie X P, Tian E G. Security consensus control for multi-agent systems under DoS attacks via reinforcement learning method. *Journal of the Franklin Institute*, 2024, **361**(1): 164–176
- 30 Feng T, Zhang J L, Tong Y, Zhang H G. Q-learning algorithm in solving consensusability problem of discrete-time multi-agent systems. *Automatica*, 2021, **128**: Article No. 109576
- 31 Long M K, Su H S, Zeng Z G. Output-feedback global consensus of discrete-time multiagent systems subject to input saturation via Q-learning method. *IEEE Transactions on Cybernetics*, 2022, **52**(3): 1661–1670
- 32 Zhang H P, Park J H, Yue D, Xie X P. Finite-horizon optimal consensus control for unknown multiagent state-delay systems. *IEEE Transactions on Cybernetics*, 2020, **50**(2): 402–413
- 33 Liu C, Liu L. Finite-horizon robust event-triggered control for nonlinear multi-agent systems with state delay. *Neural Processing Letters*, 2023, **55**(4): 5167–5191
- 34 Guzey H M, Xu H, Sarangapani J. Neural network-based finite horizon optimal adaptive consensus control of mobile robot formations. *Optimal Control Applications and Methods*, 2016, **37**(5): 1014–1034
- 35 Yu D, Ge S S, Li D Y, Wang P. Finite-horizon robust formation-containment control of multi-agent networks with unknown dynamics. *Neurocomputing*, 2021, **458**: 403–415
- 36 Shi J, Yue D, Xie X P. Optimal leader-follower consensus for constrained-input multiagent systems with completely unknown dynamics. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 2022, **52**(2): 1182–1191
- 37 Qin J H, Li M, Shi Y, Ma Q C, Zheng W X. Optimal synchronization control of multiagent systems with input saturation via off-policy reinforcement learning. *IEEE Transactions on Neural Networks and Learning Systems*, 2019, **30**(1): 85–96
- 38 Long M K, Su H S, Wang X L, Jiang G P, Wang X F. An iterative Q-learning based global consensus of discrete-time saturated multi-agent systems. *Chaos*, 2019, **29**(10): Article No. 103127
- 39 Long M K, Su H S, Zeng Z G. Model-free algorithms for containment control of saturated discrete-time multiagent systems via Q-learning method. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 2022, **52**(2): 1308–1316
- 40 Wang B J, Xu L, Yi X L, Jia Y, Yang T. Semiglobal suboptimal output regulation for heterogeneous multi-agent systems with input saturation via adaptive dynamic programming. *IEEE Transactions on Neural Networks and Learning Systems*, 2024, **35**(3): 3242–3250
- 41 Wang L J, Xu J H, Liu Y, Chen C L. Dynamic event-driven finite-horizon optimal consensus control for constrained multiagent systems. *IEEE Transactions on Neural Networks and Learning Systems*, 2024, **35**(11): 16167–16180
- 42 Lin Z L. *Low Gain Feedback*. London: Springer, 1999.
- 43 Calafiore G C, Possieri C. Output feedback Q-learning for linear-quadratic discrete-time finite-horizon control problems. *IEEE Transactions on Neural Networks and Learning Systems*, 2021, **32**(7): 3274–3281
- 44 Wang W, Xie X P, Feng C Y. Model-free finite-horizon optimal tracking control of discrete-time linear systems. *Applied Mathematics and Computation*, 2022, **433**: Article No. 127400
- 45 Wu H, Su H S. Discrete-time positive edge-consensus for undirected and directed nodal networks. *IEEE Transactions on Circuits and Systems II: Express Briefs*, 2018, **65**(2): 221–225
- 46 Lewis F L, Vrabie D L, Syrmos V L. *Optimal Control*. Hoboken: John Wiley & Sons, 2012.
- 47 Jiang Y, Kiumarsi B, Fan J L, Chai T Y, Li J N, Lewis F L. Optimal output regulation of linear discrete-time systems with unknown dynamics using reinforcement learning. *IEEE Transactions on Cybernetics*, 2020, **50**(7): 3147–3156
- 48 Jiang Yi, Fan Jia-Lu, Jia Yao, Chai Tian-You. Data-driven flotation process operational feedback decoupling control. *Acta Automatica Sinica*, 2019, **45**(4): 759–770
(姜艺, 范家璐, 贾瑶, 柴天佑. 数据驱动的浮选过程运行反馈解耦控制方法. *自动化学报*, 2019, **45**(4): 759–770)



王 巍 中南财经政法大学副教授。2019 年获得中国地质大学 (武汉) 控制科学与工程专业博士学位。主要研究方向为强化学习与自适应动态规划, 多智能体系统, 有限时域最优控制。E-mail: imagef@zuel.edu.cn

(**WANG Wei** Associate professor at Zhongnan University of Economics and Law. He received his Ph.D. degree in control science and engineering from China University of Geosciences (Wuhan) in 2019. His research interest covers reinforcement learning and adaptive dynamic programming, multi-agent systems and finite-horizon optimal control.)



王 珂 天津大学助理研究员。2023 年获得天津大学控制科学与工程专业博士学位。主要研究方向为强化学习与自适应动态规划, 微分博弈与应用, 事件触发方法。

E-mail: walker_wang@tju.edu.cn
(**WANG Ke** Assistant researcher at Tianjin University. He received his Ph.D. degree in control science and engineering from Tianjin University in 2023. His research interest covers reinforcement learning and adaptive dynamic programming, differential games and applications, and event-triggered methods.)



黄自鑫 武汉工程大学副教授. 2020 年获得中国地质大学 (武汉) 控制科学与工程专业博士学位. 主要研究方向为软体机器人, 强化学习.

E-mail: huangzx@wit.edu.cn

(HUANG Zi-Xin Associate professor at Wuhan Institute of Tech-

nology. He received his Ph.D. degree in control science and engineering from China University of Geosciences (Wuhan) in 2020. His research interest covers soft robotics and reinforcement learning.)



王乐君 重庆邮电大学讲师. 2022 年获得中国地质大学 (武汉) 控制科学与工程专业博士学位. 主要研究方向为机器人智能控制技术.

E-mail: wanglj@cqupt.edu.cn

(WANG Le-Jun Lecturer at Chongqing University of Posts and Tele-

communications. He received his Ph.D. degree in control science and engineering from China University of Geosciences (Wuhan) in 2022. His main research interest is intelligent control technology of robot.)



穆朝絮 天津大学教授. 2012 年获得东南大学控制科学与工程专业博士学位. 主要研究方向为强化学习, 自适应学习系统, 无人优化与控制. 本文通信作者.

E-mail: cxmu@tju.edu.cn

(MU Chao-Xu Professor at Tianjin University. She received her Ph.D. degree in control science and engineering from Southeast University in 2012. Her research interest covers reinforcement learning, adaptive learning systems and unmanned optimization and control. Corresponding author of this paper.)